



Simplification of Water Distribution Network Simulation by Topological Clustering – Investigation of its Potential Use in Copenhagen's Water Supply Monitoring and Contamination Contingency Plans

Kirstein, Jonas Kjeld; Albrechtsen, Hans-Jørgen; Rygaard, Martin

Published in:
Procedia Engineering

Link to article, DOI:
[10.1016/j.proeng.2014.11.248](https://doi.org/10.1016/j.proeng.2014.11.248)

Publication date:
2014

Document Version
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

Citation (APA):
Kirstein, J. K., Albrechtsen, H-J., & Rygaard, M. (2014). Simplification of Water Distribution Network Simulation by Topological Clustering – Investigation of its Potential Use in Copenhagen's Water Supply Monitoring and Contamination Contingency Plans. *Procedia Engineering*, 89, 1184-1191.
<https://doi.org/10.1016/j.proeng.2014.11.248>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

16th Conference on Water Distribution System Analysis, WDSA 2014

Simplification of Water Distribution Network Simulation by Topological Clustering - Investigation of its Potential Use in Copenhagen's Water Supply Monitoring and Contamination Contingency Plans

J.K. Kirstein^{a,*}, H.-J. Albrechtsen^a, M. Rygaard^a^a*Department of Environmental Engineering, Technical University of Denmark, Miljøvej 113, 2800 Kgs. Lyngby, Denmark*

Abstract

Topological clustering was investigated to simplify a complex water distribution network of Copenhagen, Denmark, into recognizable water movement patterns. This made it possible to assess the general transport of the water and to suggest strategic sampling locations. Through a topological analysis, the network model was divided into strongly and weakly connected clusters within selected time periods. Steady connected clusters were found by conducting a cluster analysis over all chosen selected time periods. We identified sampling locations with steady hydraulic conditions, increasing the samples' comparability over time, and locations, where samples represent the distributed and consumed water in the Nørrebro district.

© 2014 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/3.0/>).

Peer-review under responsibility of the Organizing Committee of WDSA 2014

Keywords: Clustering, Water distribution network simplification, Monitoring, Sampling, Graph theory

1. Introduction

To ensure a safe and adequate drinking water quality at consumers' taps, frequent and regular monitoring of the water quality in water distribution networks (WDNs) is needed. For this purpose, current legislation in the European Union sets several parametric values for the quality of the distributed water and the quantity of samples taken within the water supply system, although requirements for the location of sampling locations within the WDN are not stated clearly [1]. For practical reasons, authorities and utilities can only employ a limited number of samples, and it is impossible to check the water quality at every node in the pipe system. To minimize public health impacts, it is therefore important to optimize monitoring practices and to achieve the best possible coverage, given a limited number of samples available. However, the growing size of urban WDNs increases the difficulty of identifying a limited number of strategic sampling locations, where the consumers water quality and the detection of possible contaminations is covered.

* Corresponding author. Tel.: +45-60551150.

E-mail address: s103358@student.dtu.dk

If a sample indicates poor water quality or even contamination, it is not only crucial to map the distribution of the contamination, but also to identify where in the WDN the water might come from. In general, when the size and complexity of urban WDNs increases, the assessment of water movement patterns in the system becomes more and more difficult. Perelman and Ostfeld [2] approached this problem by developing a methodology that simplifies a WDN through a topological analysis. Based on graph theory, they divided a WDN into “strongly connected clusters” (SCCs) with the depth first search algorithm [3] and “weakly connected clusters” (WCCs) with the breadth first search algorithm [4]. A SCC is formed for all nodes u and v , where a directed path (sequence of distinct nodes) from u to v and a directed path from v to u exists. Where only one directed path exists, either from u to v or from v to u , a WCC is formed. With this approach, a ranked connectivity matrix was established, which could be used in case of a contaminant intrusion. To minimize the number of generated clusters in larger WDNs, Perelman and Ostfeld [5] advanced their work and proposed a methodology for how smaller clusters could be merged, e.g. by setting upper and lower bounds for when WCCs were classified as single WCCs or when they should be merged with adjacent clusters.

In the original method, SCC boundary nodes or starting points, such as service reservoirs or tanks, serve as root nodes for the identification of WCCs [2,5]. This approach risks ambiguity in terms of whether one or two WCCs are identified in the case that a SCC has two boundary nodes but only one of them has a direct downstream connection to the other node. In this instance the identification of one or two WCCs depends on the starting point of the BFS algorithm. Moreover, it was concluded [2] that the optimal starting time of the simulation and duration of the analysed time period are problem-dependent, because they influence the type and size of the clusters and connections between the generated clusters. Thus, as a novel development on previous work in [2,5], our aim was to advance and investigate a modified method by merging adjacent WCCs to generate unique clusters without ambiguity of cluster formation and consequently, no WCCs will be merged with SCCs in this study. Since clustering time duration is assumed to have a great influence on the number of generated cluster [5], we wanted to explore how varying diurnal flow regimes affect the cluster formation by analyzing for selected time periods with more or less constant flow directions. The selected time periods will be used to analyse whether recurring patterns of clusters exist over the selected periods, i.e. an identification of steady clusters. Other aims were to identify strategic sampling locations, i.e. locations where a limited number of samples has optimal coverage of the distributed water and/or can assist eventual tracking of detected contaminations; and to analyse if the approach could be useful in a contamination event by the use of mesh diagrams (i.e. cluster topology charts) of clusters will be generated, visualizing the flow of the water through the district. Copenhagen’s WDN model was used as an example network with the overall objective to simplify the network and to analyse the general transport of the water through the WDN, although all cluster results will be assessed in detail only for the sub-district Nørrebro due to the complexity of the network.

2. Methodology

The methodology of topological clustering is based on fundamental notions of graph theory [2,5], however, the methodology is further developed in our study. First, all SCCs are identified. Second, unique WCCs are identified by connecting all adjacent weakly connected compartments (no upper or lower bounds for the size of WCCs were set). For this purpose, the adjacency matrix being used for the identification of SCCs was modified by deleting all strong connections so that only weak connections are left in the matrix. This modified adjacency matrix is then added to its transposed, and the depth first search algorithm is now capable of identifying all WCCs (disguised as SCCs).

2.1. Time period selection

Cluster results from long periods could be misinterpreted because a large SCC is generated when the flow direction changes in a single pipe for a couple of minutes. Even though this issue is not eliminated by analysis within selected time periods, the number of misinterpretations is reduced by conducting the analysis over selected separate, successive time periods, because the individual results for each time period show more exact “flow trends”.

An example of how clusters resulting from analyzing a contamination case can be misinterpreted is shown in Fig. 1, illustrating a WDN with a cyclic 24-hour flow pattern. A contamination is detected in node a between 00 and 12 hours (Fig. 1a). Between 12 and 24 hours, flow directions change and the contamination can spread from node a towards node b (Fig. 1b). In both selected periods, only the formation of WCCs is observed. However, if a cluster

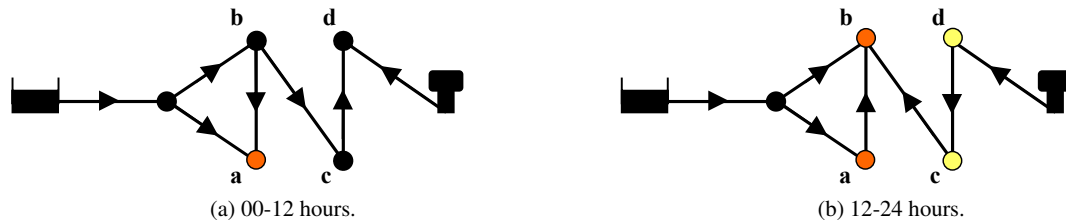


Fig. 1. Example of clustering and contamination detection at two selected time periods (color coding): (a) contamination detected at node *a* and (b) contamination detected at node *a* and *b*. Within the respective time periods, weakly connected clusters are formed. Conducting a cluster analysis from 00-24 hours, node *a* to *d* form a strongly connected cluster.

analysis is conducted from 00 to 24 hours, node *a* to *d* would form a SCC. This implies, according to the theory behind SCCs that the contamination could spread to all nodes within the cluster. In theory, however, the contamination could potentially first be spread towards node *c* and *d* after the 24 hours. Thus, if sampling is conducted on a weekly basis, for instance, the SCC generated over 24 hours has to be taken into account when using the results in terms of tracking potential transport of contaminants. Results based on cluster formation within selected periods are preferable where hourly or more frequent sensor results are available.

If a time period is chosen that covers the maximum number of flow directions (for example between 00 and 24 hours in Fig. 1), it is unclear how frequent the flow direction changes in a generated cluster. In contrast, if two selected time periods are chosen and a SCC is observed in both periods, it is more likely that the water tends to reverse frequently. To allow for this more detailed analysis, we modified the method to investigate selected separate, successive time periods. The selected time periods represent those where flow directions are assumed to be relatively constant and likely to be predicted. For the determination of periods with constant flow directions, the emptying and filling of tanks or a varying demand in the system can serve as a starting point.

2.2. Steady connected clusters

Analysing selected time periods with different flow directions, it is possible to assess the steadiness of the results by checking the clusters at different time periods for intersections of the same type of cluster. If intersections exist, overlapping in all time periods, these sections of the clusters are less time-dependent and simplify the results from various time periods with no loss of information. These clusters are referred to as steady connected clusters, either as steady WCCs (SWCCs) or as steady SCCs (SSCCs). As an example on finding steady connected clusters, a cluster analysis was conducted in two time periods (Fig. 2). During the two periods, two and four clusters were identified (Fig. 2a and b respectively). The same types of clusters were checked for intersections and two steady WCCs and one steady SCC were found (Fig. 2c). If, instead, the analysis in this example had been conducted over 24 hours, the same WCCs as in Fig. 2c would have been found. However, this is not the case for the identified SCCs. In this case SCC2, as shown in Fig. 2b, would have been found, but the information that the nodes are part of WCC1 between 00 and 12 hours is lost.

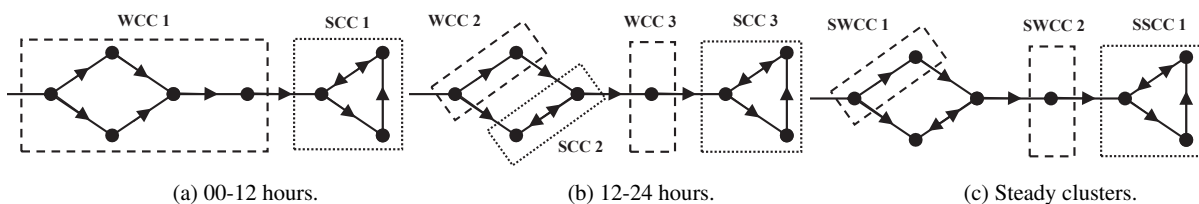


Fig. 2. Example of steady cluster analysis in two selected time periods. Clusters formed between (a) 00 and 12 hours; (b) 12 and 24 hours and (c) steady clusters formed within both periods. WCC = weakly connected cluster, SCC = strongly connected cluster; SWCC = steady WCC, SSCC = steady SCC.

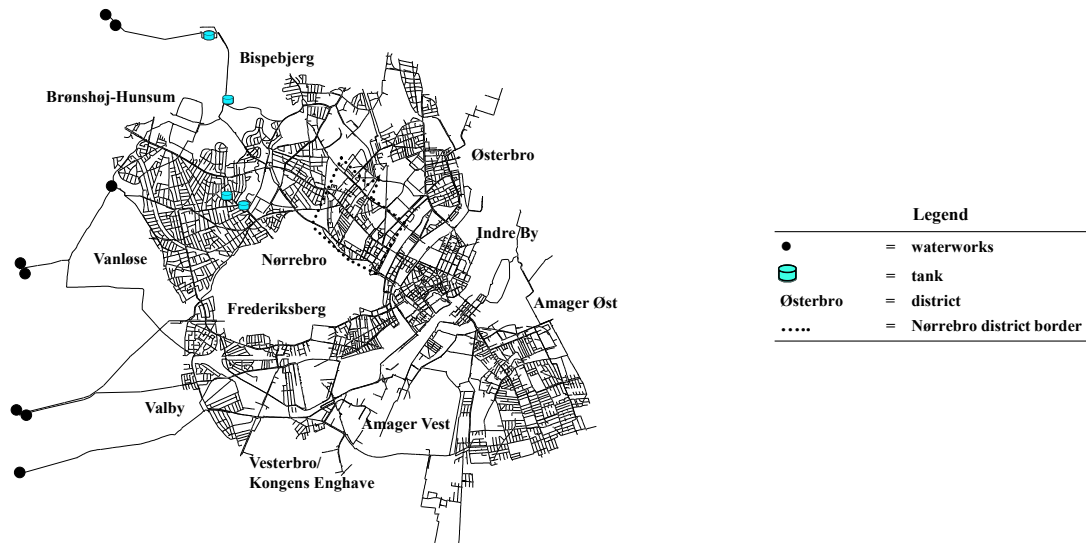


Fig. 3. Copenhagen water distribution network [6,7].

2.3. Investigated network model

In this paper, the methodology is applied to an EPANET WDN model of Copenhagen build in MIKE URBAN [6–8] (Fig. 3). The model consists of 12447 nodes, 14553 links, 8 waterworks and 4 tank systems. Only one tank system is of major importance, the Tinghøj tank system, consisting of 10 tanks in operation (most northern tank, Fig. 3) [9]. Copenhagen's WDN model simulates a weekly pattern of water consumption [10]. In total, 8758 nodes have assigned a 24-hour cyclic demand pattern, and more than 99% of all nodes are assigned to the same pattern. To provide further insight on the cluster formation and its possible implications, a more detailed analysis of the results will be conducted for the district of Nørrebro, which is located in the central part of Copenhagen (Fig. 3). A mesh diagram, i.e. clusters topology chart, will be generated to illustrate how the water in Nørrebro moves and if the cluster results can be used in terms of determining favourable sampling locations. Generally, a strategic sampling location can have two main objectives: one to cover as many nodes upstream as possible, and the other to represent the water distributed further downstream.

3. Results & Discussion

3.1. Time period selection

Between 07 and 22 hours, there was a demand above the daily average (demand multiplier > 1). In addition, the Tinghøj tank system was emptied between 07:30 and 23 hours. Due to this, two selected time periods dependent on high and low water consumption were chosen: from 00 to 07 hours (low water consumption) and from 07 to 22 hours (high water consumption). We ran the simulation for two days with the same time periods and thus a total of four selected time periods was analysed for cluster generations.

3.2. Cluster formations: Copenhagen

Cluster results for the four selected periods are listed in Table 1. Cluster results in terms of the size and type of the clusters varied greatly between the two time periods, 00-07 and 07-22 hours. However, there was less than 12% change in the number of identified clusters from day 1 to day 2 (Table 1). In the following we will focus on the results from day 1 (Fig. 4).

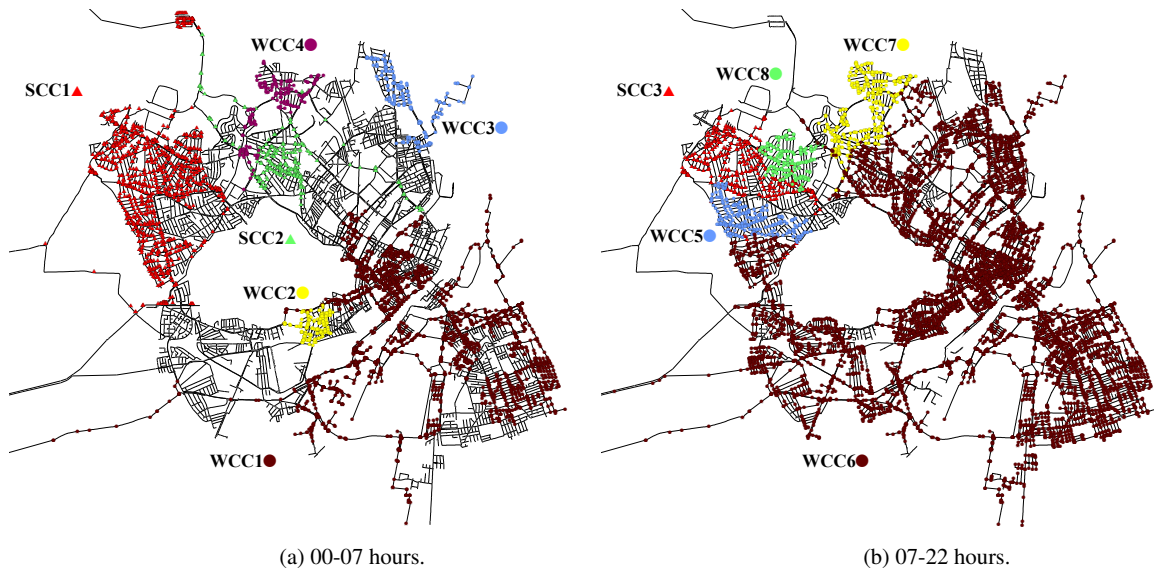


Fig. 4. Cluster formation on day 1 (color coding). For simplicity, only nodes in clusters with at least 200 nodes are shown. (a) 2 SCCs and 4 WCCs; (b) 1 SCC and 4 WCCs. SCC = strongly connected cluster; WCC = weakly connected cluster

Table 1. Cluster formations in Copenhagen's Water Distribution Network on day 1 and 2. SCC = strongly connected cluster; WCC = weakly connected cluster. A blind end is a WCC consisting of a single node.

Time period (day)	SCCs (nodes)	SCC density	WCCs (nodes)	WCC density	Blind ends
00-07 (1)	558 (4462)	8.0	411 (7559)	18.4	426
07-22 (1)	435 (1961)	4.5	137 (10317)	75.3	169
00-07 (2)	563 (4373)	7.8	400 (7659)	19.2	414
07-22 (2)	384 (1799)	4.6	128 (10496)	82.0	153

Between 00 and 07 hours, the Tinghøj tank system was being filled and a low consumption dominated the flow in the system. In this period, the demand decreased to negligible values in some nodes, and changes in the flow direction occur frequently compared to times of high water consumption. Frequent flow changes were confirmed by a relatively high number of SCCs and a higher density (nodes per cluster) of SCCs were established in the first time period (00-07 hours) than in the second (07-22). The two largest SCCs were located in the north-west, covering more than 1500 nodes, probably because the only pumping system in the network was located within SCC1 (Fig. 4a). At night, during filling of the Tinghøj tank system, the pumping influenced flow directions in areas closely located to the system. Since the Tinghøj tank system was directly connected to both SCCs, it can be assumed that water from both areas was transported into the tank system. The largest WCC (WCC1), stretched from the south-west to the eastern part of Copenhagen, covering 2770 nodes. The flow direction between all these nodes was constant for the analysed time period and no reverse flow changes were observed. Thus, it could be assumed that water from the most southern located waterworks was transported as far as the districts of Nørrebro and Amager Øst. An explanation could be that the area is located too far from the Tinghøj tank system to be influenced by it.

Between 07 and 22 hours, cluster formations changed greatly compared to the earlier time period and the water was mainly "pushed" through the system because of a higher water consumption. Thus, WCCs grew in size and dominated flow dynamics in the network (Table 1). During this period of high consumption, WCC6 extended over the entire eastern part of the network, covering more than 8000 nodes (Fig. 4b). It appeared that the flow direction of this cluster was towards southern and eastern Copenhagen. However, due to the large area covered by WCC6, the methodology tends to oversimplify the results in this time period. In addition, only one large SCC was formed in the second period (SCC3), covering a smaller area than SCC1 from the first time period.

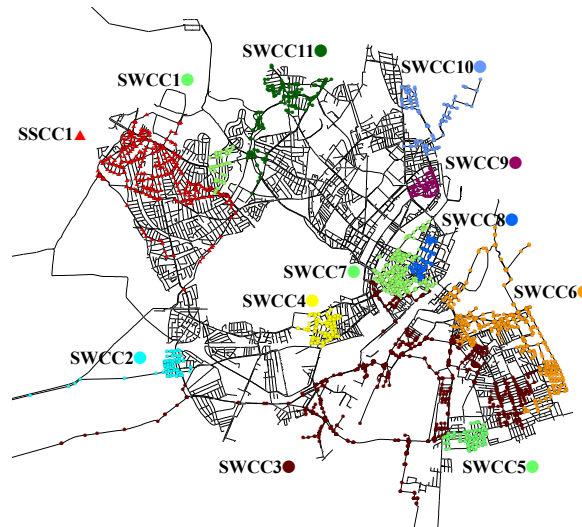


Fig. 5. Steady cluster formations with at least 200 nodes per cluster (color coding). SCC = strongly connected cluster; WCC = weakly connected cluster; SSCC = steady SCC; SWCC = steady WCC.

Table 2. Steady cluster formations in Copenhagen's Water Distribution Network. SCC = strongly connected cluster; WCC = weakly connected cluster.

Steady SCCs (nodes)	Steady SCCs density	Steady WCCs (nodes)	Steady WCCs density	Nodes not part of steady cluster
180 (1046)	5.8	499 (6667)	13.4	4734

3.3. Steady cluster formations: Copenhagen

A steady cluster analysis was conducted for the results of the four time periods. Steady cluster results are listed in Table 2 and the steady clusters with at least 200 nodes are shown in Fig. 5. Approximately 8% of all nodes in the network were part of a steady SCC. In contrast, about 54% were part of a steady WCC. Constant hydraulic conditions were observed in the south of Copenhagen; for example, steady WCC3 stretched from the south-west to the south-east of Copenhagen and covered more than 800 nodes (Fig. 5). Samples from a fixed sampling location within this cluster would be readily comparable due to the origin and distribution of water. Moreover, the greatest steady SCC was found in the north-east (steady SCC1). It can be stated that the water in this area, covering more than 500 nodes, tends to move frequently “back and forth”. Such a steady SCC is especially vulnerable because of an increased risk of spreading contamination. A steady SCC is particularly relevant for sampling, since the whole cluster is an inter-looped system and any contamination could potentially spread to all nodes after a certain time. Thus, there may be no preferable node for taking samples in a steady SCC.

More than a third of all nodes in the network were not part of a steady cluster, which means that the water origin changed between the four periods. This is particularly interesting if a fixed sampling location was to be placed in such a node, because the timing of sampling must be taken into account when assessing the results of the samples. A sample could, for example, be taken at the location every day at 13 hours and samples showing no water quality deterioration could be misinterpreted as a sign of safety. However, the water quality could differ greatly when the node was part of another cluster. Thus, if samples are taken at different points in time in non-steady clusters, it is not clear whether samples are comparable due to the potentially varying origin and distribution of the sampled water.

3.4. Limitation of clusters and their connections

The flow velocity determining whether water can be spread from one node to an adjacent node, was not taken into account. The velocity in some pipes might be so slow that the water never passes on to the extent of the identified

cluster. If one of several generated flow values in a pipe indicates a change in the flow direction, even with an infinitesimal velocity, a pipe is marked as having both flow directions. To dampen such effects of minor changes in the flow direction, our code used average flow on an hourly basis. However, this approach does not allow the conclusion that tracing flow from a node in a SCC will reach all nodes within the cluster, because connections in reality are limited by the actual distance traveled by the water. Extending the code to include the transport distance of the water in the pipes will be of great value in ensuring a better interpretation of possible monitoring coverage of identified clusters.

3.5. Sampling locations: Nørrebro

When only considering the district of Nørrebro, the general cluster results differed considerably in both analysed time periods, as for the entire WDN. All cluster connections between 00 and 07 hours are illustrated in the mesh diagram in Fig. 6. Between 00 and 07 hours, cluster No. 1 was the most upstream located cluster in Nørrebro. Thus, this cluster is a location from where several other clusters received their water. A sample indicating no contamination at 07 hours in cluster No. 1 can lead to the conclusion that most of the distributed water between 00 and 07 hours in the district is safe. Since water is distributed from cluster No. 1, which is part of WCC1 in Fig. 4, the mesh diagram confirmed the assumption and observations from WCC1 (Fig. 4) or SWCC2 (Fig. 5) that the water from the most southerly waterworks (Fig. 3) is transported as far as the district of Nørrebro. If samples should represent several other clusters farther downstream, the results identified cluster 6 and 7 as preferred locations, because of several inflows from other upstream clusters and the potential to sample water distributed to other districts (clusters No. 6 and 7 are marked with an inflow and/or outflow). Another sampling objective could be to identify a location where several sources are mixed. In this case, clusters No. 4 and 5 could be considered, since these have several inflows and outflows from different clusters. It must be noted that a sampling location has to be reconsidered completely different in the next period, because the general flow direction in the district here changed from north to south (results not shown). A similar mesh diagram for the second time period includes 751 nodes of WCC2. Thus, most of the district is part of the same WCC and in this case particular nodes could be pointed out, from where more than 70% of all nodes in Nørrebro were reached.

3.6. Mesh diagram as a contingency response tool: Contamination case in Nørrebro

Cluster No. 2 contains one of the fixed sampling locations for monitoring the water quality in Copenhagen (Fig. 6). In 2011, samples from this location proved that the drinking water was contaminated. The potential origin of the contamination was found to be within the range of four nodes, all marked with an asterisk (*) (Fig. 6). In a contingency

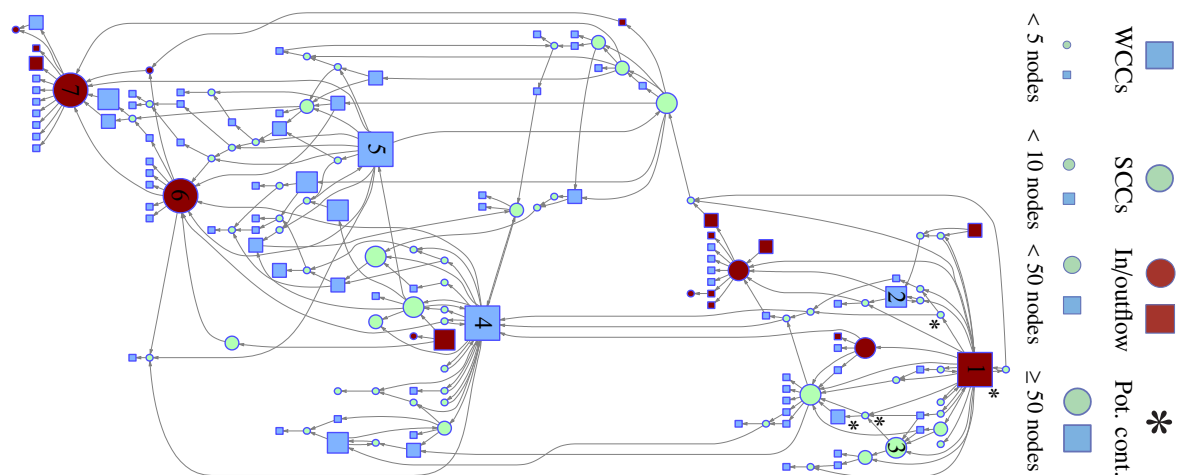


Fig. 6. Mesh diagram for day 1, Nørrebro, time period: 00-07 hours. A “*” illustrates the location of a potential contamination. Clusters marked with red have an in and/or outflow connection from the district. SCC = strongly connected cluster; WCC = weakly connected cluster.

situation the mesh diagram can be used for two objectives: narrowing the area affected by the contamination and tracking the source of the contamination. Following all outgoing connections from cluster No. 2, the mesh diagram could be interpreted as a worst-case scenario map in the selected time period, showing where the contamination could have been distributed to. This could be particularly useful in terms of immediate contamination intrusion responses for the responsible authorities. Assuming the contamination event started at 00 hours and the sample was taken before 07 hours, several clusters (such as cluster No. 3) can be excluded as potential origin due to the lack of a direct connection to cluster No. 1. However, the water in cluster No. 2 has the potential to originate from nine different clusters (Fig. 6). Of these, cluster No. 1 is the only one of the nine marked with an asterisk. Thus, cluster No. 1 is the only cluster from where the contamination could have reached the sampling location within the selected time period.

4. Conclusion

We have applied a novel methodology based on topological clustering for a complex water distribution network model with varying flow patterns and have shown the potential of analysing clusters within selected time periods. With this, we could:

- assess the difference in formation of clusters in selected time periods
- identify steady connected clusters, revealing areas of constant hydraulic conditions
- identify selected time periods where the applied methodology results in oversimplification
- create mesh diagrams showing a contamination worst-case scenario of water movement within selected time periods, and how these could be used as immediate response tools in case of contaminant intrusions

Specifically for Copenhagen's WDN model our results:

- showed that more than a third of all nodes in the network were not assigned to a steady connected cluster, indicating that water samples at various points in time vary in terms of origin and distribution
- identified several steady strongly and steady weakly connected clusters, some covering more than 500 nodes
- could be used to suggest specific sampling locations where samples will represent mixed, distributed and consumed water

Finally, we have identified a number of potential methodological improvements of the topological clustering method, of which taking flow velocities and specific connections between the clusters into consideration are especially important challenges. This is particularly important where clustering results are used for identifying contaminated nodes and backtracking the source of contamination.

References

- [1] European Parliament and Council, Directive 98/83/EC of 3 November 1998 on the quality of water for human consumption, Official Journal of the European Communities L330 (1998) 32–54.
- [2] L. Perelman, A. Ostfeld, Topological clustering for water distribution systems analysis, *Environmental Modelling & Software* 26 (2011) 969–972.
- [3] R. Tarjan, Depth-First Search and Linear Graph Algorithms, *SIAM Journal on Computing* 1 (1972) 146–160.
- [4] I. Pohl, Bi-directional and heuristic search in path problems, Doctoral thesis, Stanford University, 1969.
- [5] L. Perelman, A. Ostfeld, Water-Distribution Systems Simplifications through Clustering, *Journal of Water Resources Planning and Management* 138 (2012) 218–229.
- [6] Greater Copenhagen Utility (HOFOR), EPANET model of Copenhagen, version 2009, modified 2012, 2012.
- [7] DHI, MIKE URBAN, GIS-based urban modeling system for water distribution systems and wastewater collection systems, 2014. URL: <http://releasenotes.dhigroup.com/2014/MIKEURBANrelinf.htm>.
- [8] Environmental Protection Agency, EPANET, water distribution system modeling software, 2014. URL: <http://www.epa.gov/nrmrl/wswrd/dw/epanet.html>.
- [9] DHI, Personal communication with Henrik Sandgård Andersen from DHI, 2013.
- [10] Greater Copenhagen Utility (HOFOR), Personal Communication with Søren Lind and Charlotten Hansen from Greater Copenhagen Utility, 2013.